CENTER FOR DEMOCRACY & TECHNOLOGY

Statement of Eric Null, Director of Privacy & Data Project at the Center for Democracy & Technology, Before the California Privacy Protection Agency on Data Minimization and Use Limitations

May 5, 2022

Thank you for allowing me to speak to you today on data minimization and use or purpose limitations. I am Eric Null, the Director of the Privacy & Data Project at the Center for Democracy & Technology, a DC-based nonprofit, nonpartisan organization that is committed to protecting privacy as a fundamental human and civil right.

Data minimization and use limitations are critical data protection principles that are often overlooked and not taken seriously in the United States. Many businesses set their own data agendas, crafting essentially limitless practices in dense privacy policies. Businesses often do not think critically about their data practices nor try to limit the potential data-related harm they can cause.

Data is a commodity prone to over-collection. A survey of industry leaders in the US showed that 36% of them believe over ¾ of their data is "dark" (which is unused data and is sometimes unknown) and 63% of them believe over 50% of their data is dark.[1] A recently-leaked document from Facebook shows the company "has no idea where all of its user data goes, or what it's doing with it," which would make it difficult to comply with the EU's General Data Protection Regulation own data minimization and purpose limit requirements.[2] And one broader EU study showed 72% of companies collected data they ended up not using.[3]

Anecdotal examples exist, too. Mobile apps, like Angry Birds and the infamous Brightest Flashlight App, have had a history of collecting location data without legitimate purpose.[4] Data brokers, who exist in significant part because of data over-collection and -retention, have in particular capitalized on this trend. Just this week we saw reports of a data broker selling location data of people who visited Planned Parenthood clinics that the broker collected by using software development kits from various mobile apps that track location for who knows what, if any, purpose.[5] We learned today that one data

[1] https://www.splunk.com/pdfs/dark-data/the-state-of-dark-data-report.pdf at 7.
[2] https://www.vice.com/en/article/akvmke/facebook-doesnt-know-what-it-does-with-your-data-or-where-it-goes.
[3] https://info.purestorage.com/rs/225-USM-292/images/Big%20Data%27s%20Big%20Failure_UK%281%29.pdf?aliId=64921319.
[4] https://www.nbcnews.com/technolog/shock-dark-flashlight-app-tracks-your-location-1B7991120.
[5] https://www.vice.com/en/article/m7vzjb/location-data-abortion-clinics-safegraph-planned-parenthood.

broker even made that same location data available for free.[6] Also, a few years ago, mobile carriers were caught providing cell-site location data to third party data brokers that ended up in the hands of bounty hunters.[7]

For their part, *people don't want* companies to collect such extensive data about them. A 2020 survey showed that almost 80% of Americans expressed concern over sharing personal information with online businesses.[8] In 2019, a significant majority of Pew survey respondents were concerned about how much data about them is collected by businesses, and similar numbers believed the risks to such data collection outweighed the benefits.[9]

Data minimization and use limitations are potential solutions to those problems. At its strictest, the minimization principle requires companies to collect only the data they need to provide the product or service, and nothing else.[10] Many definitions, including California's, are broader and tie minimization to specific uses.

These are important substantive provisions in the CPRA. Your agency should engage meaningfully with the plethora of uses for which companies collect data and decide whether there are harmful uses that require curtailing.

One approach taken by CDT in its comprehensive privacy framework is to prohibit certain harmful data processing practices when those practices are not required to provide or do not add to the functionality of the product, service, or specific feature that a person has requested. Those practices include

- Biometric tracking,
- Precise location tracking,
- Cross-device tracking,
- Tracking of children under 13 years of age,
- The content of and parties to communications,
- Audio and visual recording, and
- Health information.

---

[6] https://www.vice.com/en/article/g5qaq3/location-data-firm-heat-maps-planned-parenthood-abortion-clinics-placer-ai.
[7] https://www.vice.com/en/article/43z3dn/hundreds-bounty-hunters-att-tmobile-sprint-customer-location-data-years.
[8] https://winmr.com/global-crisis-in-trust-over-personal-data.
[9] https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information.
[10] https://www.accessnow.org/cms/assets/uploads/2021/05/Data-Minimization-Report.pdf.

These uses (when employed beyond the functionality of the product or service) cause harm without countervailing benefits, and they should be limited.

In addition to that list, your agency should limit secondary data use. The CPRA states that companies can collect data that is "reasonably necessary and proportionate to achieve" the original purpose of the collection or "another disclosed purpose that is compatible with the context in which the personal information was collected." This language makes clear the importance of disclosing essentially all uses, thus disallowing most secondary uses already. Any allowed secondary uses are limited to only those uses compatible with the context of the original collection, meaning the secondary purpose should be directly connected to the original purpose. For instance, if a business collects a person's phone number for account verification purposes, it could not then later use that data to serve ads because that is a wholly different context than the original collection.[11]

Your agency should also limit discriminatory data use. We know that data can be used to discriminate both directly and through algorithmic discrimination.[12] Years ago, the U.S. Department of Housing and Urban Development sued Facebook for letting housing advertisers filter out users on the basis of their race, color, religion, sex, familial status, nationality, or disability.[13] Amazon previously used an HR recruiting tool that downgraded women on the basis of their gender because Amazon's training set for the software included resumes from mostly men.[14] Under no circumstances should companies be allowed to use data, or train algorithms, in ways that discriminate against people based on protected characteristics, particularly in housing, credit, employment, insurance, and education.

One final note on form – we all know privacy policies are poor vehicles for informing people about actual data practices. People don't read them, they're too long and difficult to read, and even those who do read them will find a confusing laundry list of practices a business "may" engage in. Without describing actual practices, it is impossible to understand what data businesses have about people and how it is used. The agency could clarify that businesses should create easy-to-read summaries that describe the most salient data practices that businesses actually engage in.

Thank you for the chance to speak to you today, I look forward to working with the agency.

---

[11] https://www.engadget.com/2018-09-28-facebook-two-factor-phone-numbers-ads.html.
[12] https://www.lawyerscommittee.org/wp-content/uploads/2021/08/FTC-civil-rights-and-privacy-letter-Final-1.pdf.
[13] Assistant Secretary for Fair Housing & Equal Opportunity v. Facebook, Inc., Case No. 01-18-0323, Aug. 13, 2018.
[14] https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.